PROCEEDINGS, TOUGH Symposium 2012
Lawrence Berkeley National Laboratory, Berkeley, California, September 17-19, 2012

ADVANCES IN HYDROGEOPHYSICAL JOINT INVERSION

Michael Commer, Michael B. Kowalsky, Stefan Finsterle, Gregory A. Newman

Lawrence Berkeley National Laboratory
Earth Sciences Division, 1 Cyclotron Rd.
Berkeley, California, 94720, USA
e-mail: MCommer@lbl.gov

ABSTRACT

The inversion framework iTOUGH2 provides inverse modeling capabilities for TOUGH2, a general-purpose simulator for multiphase, multi-component, nonisothermal flows in multidimensional fractured-porous media. We have developed a parallel version of iTOUGH2, MPiTOUGH2, which realizes a hierarchically parallel architecture using the Message Passing Interface. This architecture allows employing large numbers of parallel processes for running many-parameter inverse problems on large meshes, such as those occurring in pixel-based parameterizations.

We have further combined the parallel hydrological inversion method with a geophysical modeling and inversion framework, called Electromagnetic Geological Mapper (EMGeo). The combined inversion method provides a tool for investigating coupled hydrogeophysical processes in the context of, for example, geologic CO_2 storage, geothermal system characterization, and environmental remediation and monitoring applications. In this work, we focus on the introduction of parallel computing paradigms to iTOUGH2, and carry out basic scalability tests. Further, exploiting the forward simulation capabilities of EMGeo, we present synthetic data inversions, demonstrating the potential model resolution enhancement obtained through joint inversion of hydrological and geophysical measurements.

INTRODUCTION

Joint inversion of hydrological and geophysical data has been recognized as a means of enhancing the typically sparse coverage of hydrological data, and thus achieving enhanced parameter resolution. A number of deterministic and stochastic joint inversion approaches have been reported. We refer the reader to the works of Kowalsky et al. (2004, 2005), and Finsterle and Kowalsky (2008), and related works referenced in those for a comprehensive overview of various methods.

Whenever a significant number of hydrological (and geophysical) simulations are carried out, such as in inverse modeling, computational efficiency becomes paramount. In this work, we advance the inverse modeling capabilities of iTOUGH2 by introducing parallel computing paradigms, using the Message Passing Interface (MPI), both into the underlying (forward) simulator given by TOUGH2, as well as the inverse modeling scheme. We further enhance the joint inversion capabilities by merging the parallel iTOUGH2 framework, called MPiTOUGH2, with the geophysical modeling methods provided by the geophysical simulator EMGeo. EMGeo is a 3D finite-difference scheme for the simulation and inversion of electrical/electromagnetic geophysical data types, including controlled-source electromagnetics, magnetotellurics, electrical resistivity tomography, and (spectral) induced polarization (Newman and Alumbaugh, 1997; Commer and Newman, 2008; Commer et. al., 2011).

In this and an accompanying work, two hydrogeophysical joint inversion approaches are used. The first strategy (presented in the second example of Kowalsky et al., 2012, this issue) involves a two-step approach. Field-scale electrical resistivity distribution maps, which would be obtained by inverting geophysical tomography data, are matched with their counterparts calculated from tracer concentrations. The latter are updated from the parameter estimates obtained during the iterative hydrological inversion procedure. The second approach, on which we focus in this work, involves a direct joint inversion of hydrological and geophysical data. After each iTOUGH2 model updating step, the subsurface electrical resistivity distribution is updated from the current hydrological parameter state. Geophysical measurements are then simulated by the modeling methods available through EMGeo and matched with field data. We focus on the geophysical resistivity tomography (ERT) method for the geophysical data component, which is sensitive to the subsurface electrical resistivity. Hence, we use Archie's law as the petrophysical model to link hydrological to electrical attributes:

$$\sigma = \sigma_f \Phi^m S^n,$$

where σ is bulk electrical conductivity, σ_f is fluid electrical conductivity, Φ is porosity, S is saturation, and m and n are site-specific parameters. The joint inversion capabilities are demonstrated by carrying out synthetic inversion studies based on experiments performed at the U.S. Department of Energy Integrated Field Research Challenge Site (IFRC) at Rifle, Colorado.

METHODOLGY

The aforementioned geophysical component EMGeo is a geophysical data inversion package by itself, using nonlinear conjugate-gradient based optimization strategies. At the current development stage, we only utilize the inverse modeling capabilities of iTOUGH2, namely the Levenberg-Marquardt modification of the Gauss-Newton algorithm (Finsterle and Kowalsky, 2011). Parallelizing multiple forward simulations is essential in solving parameter estimation problems with many unknowns. iTOUGH2 includes this feature, using the Parallel Virtual Machine (PVM) (Finsterle, 1998). However, limitations are currently imposed by the fact that each forward problem can only be solved serially, i.e., using one process.

Hierarchical Parallel Architecture

The overall parallel architecture of the hydrogeophysical inversion scheme presented here borrows a parallel layout used for the inversion of large-scale electromagnetic data sets (Commer et al., 2008), and combines it with slightly modified techniques used by the parallel simulator TOUGH-MP (Zhang et al., 2008) for the hydrological flow-simulation component. Both the fluid-flow and geophysical *forward* simulators divide their respective simulation domains into a number of subdomains. The geophysical simulator uses proprietary parallel iterative solvers for solving either Maxwell's equations, for simulating frequency-domain EM methods, or the Poisson equation, for potential field simulations (such as electrical resistivity soundings), on large structured Cartesian finite-difference grids. TOUGH-MP uses the parallel linear-equation solver AZTEC (Tuminaro et al., 1999) in conjunction with the TOUGH2 simulation framework to solve the fluid flow problem's mass and energy balance equations (Zhang et al., 2008). The underlying simulator TOUGH2 employs integral finite-differences on structured or unstructured grids for the spatial descretization, while time is discretized using standard first-order backward finite differences.

Here, we outline some new but straightforward aspects of adding another parallel level to the inverse scheme, employing the Message Passing Interface (MPI). In the inverse problem, the sensitivity calculations involved in the parameter estimation steps are by far the most computationally intense. As illustrated in Figure 1, these calculations are distributed among multiple processor groups, also called *model communicators*. Each one is a separate MPI communicator and holds a copy of the simulation domain.

In the Levenberg-Marquardt implementation of iTOUGH2, the parameter sensitivity matrix is calculated by a perturbation approach. In the parallel version MPiTOUGH2, each model communicator is in charge of a subset of the parameter space and the corresponding forward simulations of their perturbations. Note that one parameter makes one column in the sensitivity matrix. The communicators can be scaled up according to the size of the matrix system to be solved in the forward problem, using the aforementioned parallel structures of the geophysical and fluid-flow simulators. Thus, if massively parallel resources are available, inverse problems with both large simulation domains and large parameter numbers can be solved, within reasonable computing times, by adequately increasing the size and number of the model communicators.

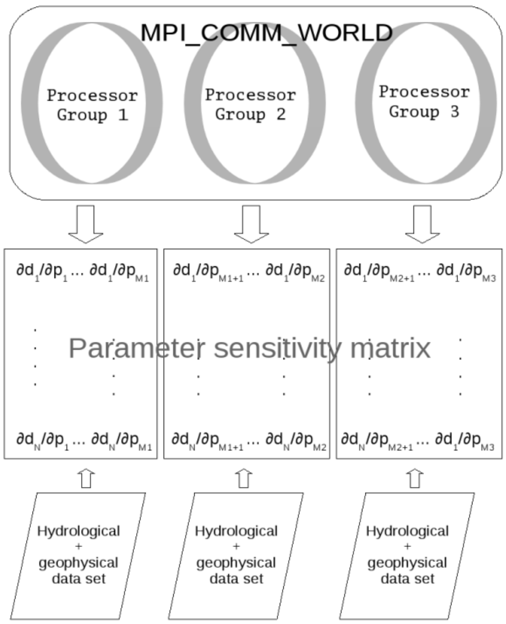

Figure 1. Illustration of the parallel layout used in the parallel inversion scheme MPiTOUGH2. Multiple forward simulations are distributed across multiple processor groups, where each group calculates a part of the parameter sensitivity matrix.

Parallel Scaling Tests

In this work, we use a 3-D TOUGH2 model created for the inversion of tracer data collected during the 2007 "Winchester" surveys (Williams et al., 2011) for some parallel scalability tests. The TOUGH2 grid comprises 52,235 elements and 151,149 connections, and has been used by Kowalsky et al. (2012) for simulating the flow cell of interest in inversions of tracer breakthrough data using a variety of hydraulic permeability parameterizations. We consider a three-parameter setup here, but for brevity, we omit further details about the model and data and refer to the aforementioned publication.

Parallel forward solution scalability test

Figure 2 shows run times for a forward simulation using from one to eight Intel® Xeon 3.20 GHz processors of a desktop computer. The bulk of the computing time is used for the iterative solution of the flow equations, including the AZTEC parallel solver's internal message passing. More (external) message passing of primary thermodynamic variables across subdomain boundaries is involved in the equation-of-state update before each Newton-Raphson iteration. However, it appears that the overhead of this message passing is negligible, owing to communication improvements reported by Zhang et al. (2008). The second-largest computing time fraction is needed for assembling the flow equation's matrix system at each Newton-Raphson iteration, followed by updating the equation-of-state and miscellaneous routines. In this example, because the mesh size is not highly demanding of the computing hardware, we observe a quick flattening of the run times when using more than three parallel tasks. Further increasing the number of parallel tasks would thus not further speed up the simulation time, due to the solver library's internal message passing overhead.

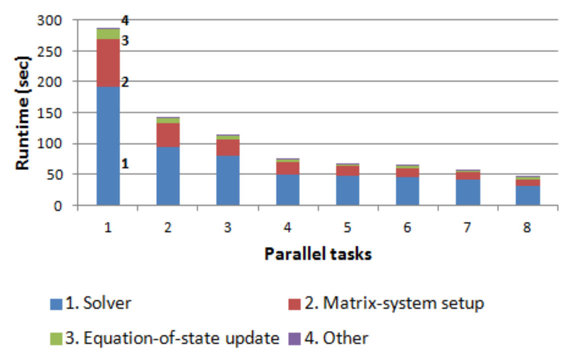

Figure 2. Runtimes for a parallel forward simulation using 1-8 parallel processes on a desktop computer.

Parallel inverse solution scalability test

The same model and computing hardware as used before is now employed in a three-parameter hydrological tracer data inversion, based on studies of optimal permeability model parameterizations (Kowalsky et al., 2012). Since the inversion involves multiple forward simulations, we expect that the behavior of the runtimes behaves similar to Figure 2. This can be observed in general in Figure 3, which depicts the average runtimes per inversion iteration. However, the runtime increases going from two to three parallel tasks. It is common in parallel numerical calculations for the solution accuracy to vary with the number of parallel tasks employed, owing to the round-off error propagation of the floating-point arithmetic (Asserrhine et al., 1995). In this particular example, it could be observed that the differences in the iterative forward solutions propagate in a way that leads the three-process inverse solution onto a solution path in model space with a lower convergence rate. Figure 4 further illustrates this effect by comparing the data objective functional values between the one- and three-process runs. While we do not elaborate further on details of this observation, we mention that both solutions yielded similar parameter estimates close to the true model. This round-off error effect may need to be taken into account as part of a systematic modeling error, especially in the presence of ill-posed inverse problems, where it is likely to be enhanced. The corresponding parameter-estimation uncertainty may then be quantified by carrying out multiple inversion realizations using different parallel configurations. Note that this can also be emulated if fewer physical processors than parallel processes are available.

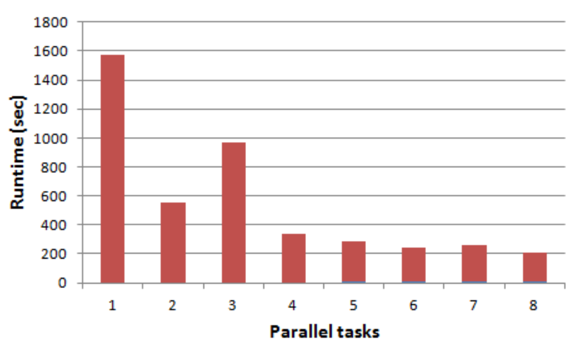

Figure 3. Runtimes for a parallel inverse solution using 1-8 parallel processes on a desktop computer.

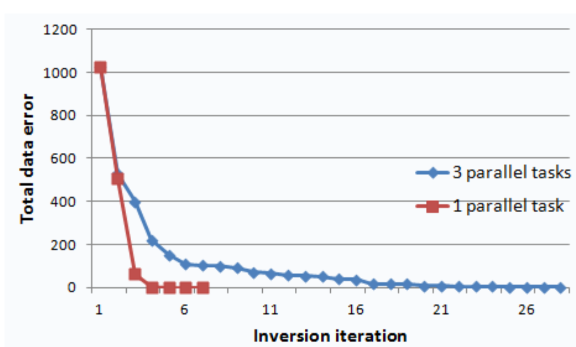

Figure 4. Data fitting error functional convergence for the one- and three-processor inversion runs of the inversion scalability test (Figure 3).

INVERSE MODELING RESULTS

The simulation cases generated for this next study are based on experiments in a shallow aquifer at the IFRC site. The aquifer is 2.5–3 m thick and is located on the floodplain of the Colorado River in alluvium, with the base given by the impermeable Wasatch formation (Williams et al., 2011). One part of the 2011/12 "Best Western" field experiments, on which we base our simulations, involved a conservative bicarbonate tracer injection to study uranium desorption in the absence of biostimulant-dependent uranium immobilization effects. The synthetic data used in the following inversion studies are generated from a subset of the survey geometries and measurement time intervals of this bicarbonate tracer injection experiment. Figure 5 illustrates the spatio-temporal layout of the simulated experiment. We consider an eight-parameter layered model, with vertically varying absolute hydraulic permeabilities as unknowns. As shown in the upper panel, the eight layers cover the aquifer's depth from 3 to 6 m below the surface. The tracer was introduced in the aquifer at the depth range indicated by the white symbols in the left part of the second panel. Groundwater flows along the *x*-axis from left to right, and was simulated for a period of 149 days after injection begin. The principal effect of the tracer is an increased flux of dissolved ions (in this case Na^+ and HCO_3^-) resulting in an electrical conductivity increase, where panels 2–6 show the reciprocal: electrical resistivity in units of Ωm.

The data set comprises a total of 219 average fluid electrical conductivity (EC) samples (given in units of μS/cm) measured in four boreholes, which are indicated by black lines in Figure 5. The EC data span a time period of more than 140 days. The data errors are given by standard deviations of 25. Using this standard deviation, which amounts to roughly 1% of the mean data amplitude, Gaussian noise with zero mean is added to the data. The geophysical data set is given by ERT measurements at 30 days. The ERT borehole array has altogether 30 electrodes distributed over four wells (indicated by red symbols). Inverted data comprise a subset of six source current electrode pairs and a total of 123 receiver electrode pairs, where one receiver electrode pair constitutes one geophysical data point. Gaussian noise with zero mean and a standard deviation of 1% of the data amplitude is added to the data, in addition to a noise floor. The ERT data are characterized by a much larger dynamic range than the range of the EC data. Thus, the used standard deviation implies a lesser degree of noise distortion than is the case for the EC data set.

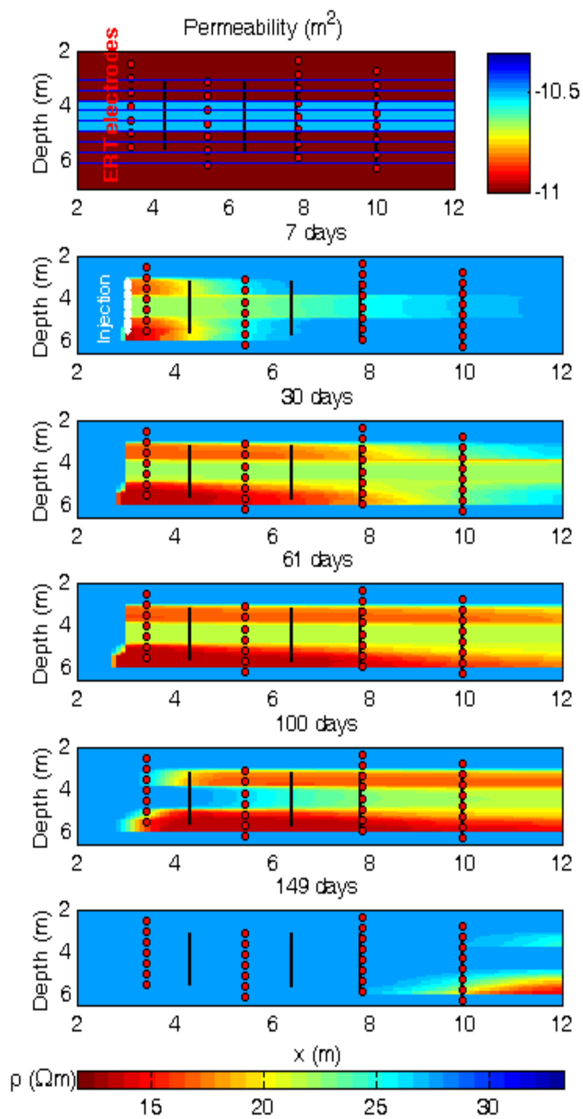

Figure 5. Spatio-temporal setup and plume development of an electrically conductive tracer injection simulated after a bicarbonate injection experiment at the IFRC site. The upper panel shows the true permeability model, where the horizontal lines indicate the boundaries of the eight layer parameters. Panels 2–6 show the change of the aquifer's electrical resistivity over a time period of 149 days after injection.

Regularization of the Inverse Problem

Many-parameter inverse problems, such as pixel-based inverse problems, are typically underdetermined, requiring the introduction of additional stabilization terms into the otherwise ill-posed problem. Two constraint types can be activated in iTOUGH2; these are *prior knowledge* and *smoothness constraints* (Finsterle, 2009). For the eight-parameter inverse problems in this work, we also observed a significant non-uniqueness problem that can be alleviated by introducing additional constraints. We introduce a practical method of automatically defining smoothness constraints in the presence of many parameters. The approach uses the ensemble of interfaces (i.e., the connection block in the TOUGH2 input) between the elements of the simulation domain. From the elements belonging to inversion parameters, all (TOUGH2) connections are determined. If there exists at least one connection between the element ensembles of two given parameters, a regularization term is introduced for this parameter pair. In our case, an element ensemble is given by a layer parameter. This idea implies that the absence of connections between two neighboring parameters would allow for a sharp (unconstrained) contrast of their estimates. The regularization is incorporated into the inverse problem by augmenting the parameter sensitivity matrix by one line per unique parameter connection.

One important aspect of constrained inversions is to properly weight the regularization terms against the information given by the calibration data. Figure 6 demonstrates the influence of the regularization weight, here called β-factor, on some trial inversions for the 8-layer test case. The β-factor is multiplied to the regularization term's objective functional, and thus has the effect of balancing it with the data-fitting component. All inversions used a homogeneous starting model with a negative log (base 10) permeability of -10.75. Each column group in Figure 6 pertains to the results for one layer parameter, where the first (left) column represents the true model, and the second, third, and fourth columns represent, respectively, the result from a regularized inversion with β=1, β=0 (no regularization), and β=10.

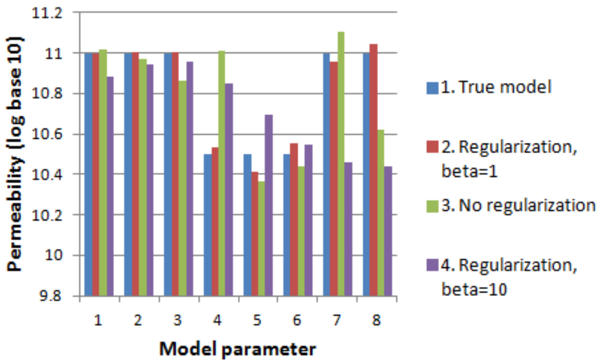

Figure 6. Inversion results with different smoothing regularization settings. Shown are permeability (negative log base 10) estimates for the 8-layer model. For each layer parameter, the results using a β-factor of 1 (red), 0 (green), and 10 (purple) are shown with the true model (blue).

While the EC data of these inversions were not contaminated with noise, their assumed error standard deviation in the inversion was 250. With these error assumptions, a regularization weight of β=1 leads to parameter estimates very close to the true model, with the corresponding data fits shown in Figure 7 (red curves). Neglecting the regularization causes an overestimated parameter 4, compensated for by an underestimated parameter 8. The resulting data fits, shown by green curves in Figure 7, are very similar to the regularized (β=1) result, indicating the non-uniqueness of the unconstrained inverse solution.

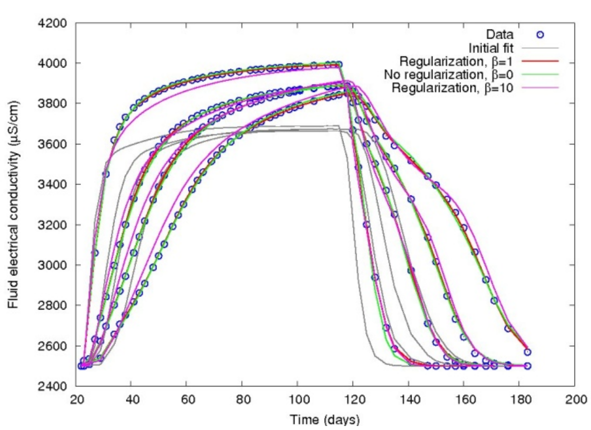

Figure 7. EC data fits calculated from inversion results with different smoothing regularization settings. The corresponding parameter estimates are shown in Figure 6.

Imposing a factor of β=10 also leads to a parameter overshoot of the low-permeability layers, where now the smoothing constraint forces the permeability to gradually decrease for the deeper layers. This happens at the expense of the data fits, as shown by the worsening fits (purple curves in Figure 7).

For all inversion runs carried out in the following, a regularization parameter of β=1 was found to be optimal.

Hydrological Inversion

In the following, all inversions are initiated with the homogeneous permeability model (log permeability of -10.75) as used before. Figure 8 shows the results of an inversion of the EC data set, where the upper part (a) shows original data (black), and initial (blue) and final (red) data fits. The model result in the lower figure (b) was obtained after 10 Levenberg-Marquardt iterations, with the same colors referring to the true, initial, and final models, where the error bars for the final model indicate two standard deviations. A good overall EC data fit is achieved by the final model. However, the noise distortion causes layer parameters 1 and 3 (from top) to deviate beyond the estimated parameter standard deviations. However, the inversion achieves an acceptable data fit, indicating that despite the regularization constraints, the non-uniqueness problem is amplified by the data's noise distortion.

Hydrogeophysical Joint Inversion

Next, we include both the geophysical ERT data and the EC data sets in the inversion. Figures 9a and 9b again show original data, and the initial and final data fits for both the ERT and EC data sets, respectively. A greatly improved model result (c) is achieved by the combined data set. The final model also improves the fit of the EC data – a 13% decrease from the total EC data fitting error in the previous EC data inversion. The total data fitting errors are plotted in Figure 10 against the inversion iterations for both inversion cases. Note that the initial total error of the EC data inversion is identical to the EC component of the joint inversion, since the homogeneous starting model produces no contribution to the error function's regularization term. We observe that the EC data errors have a similar convergence rate, which indicates that the influence of the EC and ERT components is well balanced in the joint inversion.

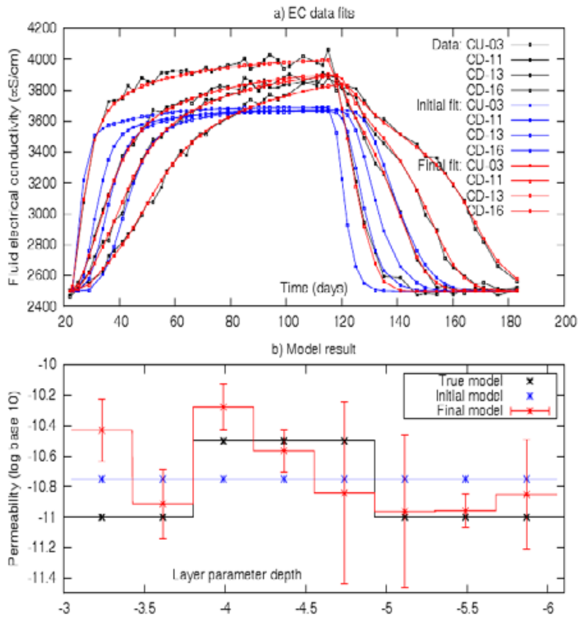

Figure 8. Inversion of the hydrological (EC) data set for the 8-layer parameter model. The boreholes named CU-03, CD-11, CD-13, and CD-16 correspond to the positions shown in Figure 5 (black lines) at x=4.3, x=6.4, x=7.8, and x=10 m, respectively.

CONCLUSIONS

We have developed a parallel framework for the joint inversion of hydrological and geophysical data. The value of geophysical data, usually covering larger survey areas than hydrological data and being cost effective, has been recognized. The development of efficient tools is thus important to fully exploit the potential resolution improvements given by combined hydrogeophysical data sets, and to investigate optimal spatiotemporal survey layouts for field applications.

Inverse modeling in general poses large computing demands. These are addressed by joining different parallelization levels, both over the forward problem, i.e., the simulation domain, and over the inverse problem, i.e., the parameter space.

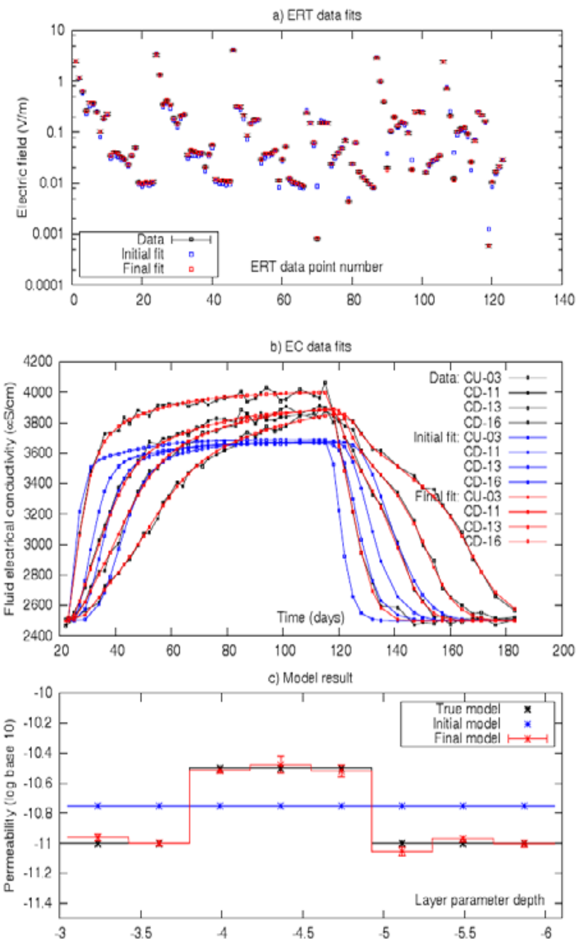

Figure 9. Joint inversion of the hydrological (EC) and geophysical (ERT) data sets for the 8-layer parameter model.

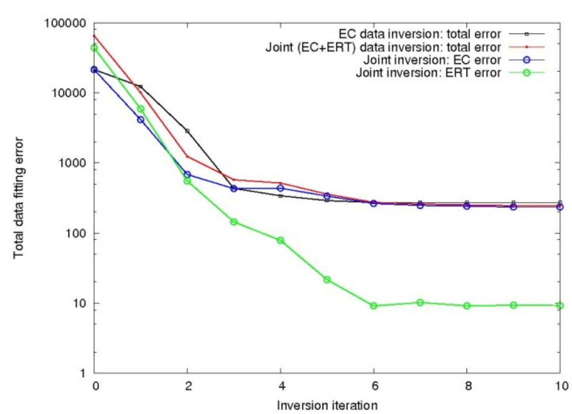

Figure 10. Comparison of total data fitting errors for the EC-data inversion (black) and joint inversion. For the joint inversion, the total (red), and separate EC (blue) and ERT (green) errors are shown

Our parallel hydrogeophysical parameter estimation framework maximizes scalability by allowing the increase of both size and number of communicators sharing the whole inverse modeling problem. The scalability of the hydrological simulator is particularly important for the joint inversion with geophysical data, because of the typical large-scale nature of geophysical inverse problems. Limitations are currently imposed by memory constraints, owing to storage requirements of a potentially large copy of the parameter sensitivity matrix on each parallel process.

The shown examples have demonstrated the potential model-resolution enhancements provided by geophysical data. It has also been shown that a properly chosen regularization parameter is essential for stabilizing the inherently ill-posed geophysical inverse problem. For the presented results, the petrophysical transform was assumed to be known. A next step will evaluate whether uncertainty in the petrophysical model can be accounted for by considering some of its parameters to be estimated in the inverse problem, utilizing the sensitivity of the geophysical data to these unknowns.

ACKNOWLEDGMENT

This work was supported by a LBNL Laboratory Directed Research and Development program granted to the authors.

REFERENCES

Asserrhine, J., J.-M. Chesneaux, and J.-L. Lamotte, Estimation of round-off errors on several computers architectures, *Journal of Universal Computer Science*, 1 (7), 454-468, 1995.

Commer, M., G.A. Newman, New advances in controlled source electromagnetic inversion, *Geophysical Journal International*, 172, 513-535, 2008.

Commer, M., G.A. Newman, J.J. Carazzone, T.A. Dickens, K.E. Green, L.A. Wahrmund, D.E. Willen, and J. Shiu, Massively-parallel electrical-conductivity imaging of hydrocarbons using the Blue Gene/L supercomputer, *IBM Journal of Research and Development*, 52-1/2, 93-103, 2008.

Commer, M., G.A. Newman, K.H. Williams, and S.S. Hubbard, Three-dimensional induced polarization data inversion for complex resistivity, *Geophysics*, 76(3), F157-171, 2011.

Finsterle, S., and M.B. Kowalsky, A truncated Levenberg-Marquardt algorithm for the calibration of highly parameterized nonlinear models, *Computers and Geosciences*, *37*, 731–738, 2011.

Finsterle, S., *Parallelization of iTOUGH2 using PVM*, Report LBNL-42261, Lawrence Berkeley National Laboratory, Berkeley, Calif., Oct. 1998.

Finsterle, S., and M.B. Kowalsky, Joint hydrological-geophysical inversion for soil structure identification, *Vadose Zone J.*, *7*, 287-293, 2008.

Finsterle, S., What's new in iTOUGH2?, Proc. TOUGH Symp., Lawrence Berkeley National Laboratory, Berkeley, Calif., 2009.

Kowalsky, M.B., S. Finsterle, and Y. Rubin, Estimating flow parameter distributions using ground-penetrating radar and hydrological measurements during transient flow in the vadose zone, *Adv. in Water Res.*, 27(6), 583-599, 2004.

Kowalsky, M.B., S. Finsterle, J. Peterson, S. Hubbard, Y. Rubin, E. Majer, A. Ward, and G. Gee, Estimation of field-scale soil hydraulic and dielectric parameters through joint inversion of GPR and hydrological data, *Water Resour. Res.*, 41, W11425, doi:10.1029/2005WR004237, 2005.

Kowalsky, M. B., S.A. Finsterle, K.H. Williams, C.J. Murray, M. Commer, D. Newcomer, A. Englert, C.I. Steefel, and S.S. Hubbard, On parameterization of the inverse problem for estimating aquifer properties using tracer data, *Water Resour. Res.*, 48, W06535, doi:10.1029/2011WR011203, 2012.

Newman, G.A., and D.L. Alumbaugh, 3-D massively parallel electromagnetic inversion - Part I. Theory, *Geophysical Journal International*, 128, 345-354, 1997.

Tuminaro, R.S., M. Heroux, S.A. Hutchinson, and J.N. Shadid, Official Aztec user's guide - version 2.1, Massively Parallel Computing Research Laboratory, Sandia National Laboratories, Albuquerque, NM, 1999.

Williams, K.H., P.E. Long, J.A. Davis, C.I. Steefel, M.J. Wilkins, A.L. N'Guessan, L. Yang, D. Newcomer, F.A. Spane, L.J. Kerkhof, L. McGuinness, R. Dayvault, and D.R. Lovely, Acetate availability and its influence on sustainable bioremediation of uranium-contaminated groundwater. *Geomicro. J.*, 28(5-6),519-539, 2011.

Zhang, K., Y.-S. Wu, and K. Pruess, *User's Guide for TOUGH-MP – A Massively Parallel Version of the TOUGH2 Code*, Report LBNL-315E, Lawrence Berkeley National Laboratory, Berkeley, Calif., May 2008.